Francis Chan

Data Analyst

330 De Neve Dr, Los Angeles, California 90024 | (408) 824-8426 | francischanziheng@gmail.com |
https://www.linkedin.com/in/francischan0806/ | https://github.com/francis008 | https://francis-chan.netlify.app/

Aspiring data analyst with a strong foundation in statistics, machine learning, and data visualization. Experienced in R, Python, SQL, and Tableau with hands-on projects involving predictive modeling, classification, and data wrangling. Passionate about turning complex data into actionable insights to support data-driven decision-making in business and technology environments.

EDUCATION

University of California, Los Angeles — **Los Angeles, CA**
Bachelor of Science in Statistics & Data Science — *Sep 2023 – Jun 2025*
De Anza College — **Cupertino, CA**
Associate of Science in Mathematics, Dean's List 2021-2023 — *Sep 2021 - June 2023*

TECHNICAL SKILLS

R, SQL, Tableau, Git, Python(pandas, numpy, scikit-learn), Stata, Java, HTML, CSS, Microsoft Excel (vlookup, pivot table)

COURSEWORKS

Math: Monte Carlo Simulation, Linear Models, Optimizations, Mathematical Statistics, Data Analysis & Regression, Computational Statistics, Multivariable Calculus, Vector Calculus, Differential Equations, Linear Algebra, Discrete
Computer Science: Data Abstraction and Structure, Statistical Programming, Data Visualization
Business: Analytics in Accounting, Microeconomics, Macroeconomics

PROJECTS

Elon Musk X Sentiment & Stock Market Volatility Prediction (Python, XGBoost, Transformer)
- Built an end-to-end machine learning pipeline that predicts short-term TSLA stock volatility from Elon Musk's tweet sentiment and engagement metrics.
- Applied multiple NLP models (VADER, FinBERT, Twitter-RoBERTa) to label tweet sentiment; engineered hybrid features and aligned them with intraday market volatility using yfinance.
- Achieved 0.85 accuracy, 0.87 precision, and 0.97 recall for high-volatility days using XGBoost, outperforming direction-based models.

NBA Game Outcome Prediction Using Statistical Modeling (Python)
-Developed a predictive analytics pipeline using logistic regression, SVM, and Random Forests to forecast NBA 2023–2024 game outcomes, achieving 69% test-set accuracy.
- Engineered matchup features (e.g., win differentials, home/away effects) and tuned hyperparameters to optimize model generalizability.
- Created a data-driven win probability framework to support strategic decision-making in sports analytics.

Fraud Detection and Transaction Analysis (XGBoost + SQL)
- Performed EDA on multi-table transaction data using advanced SQL (joins, CTEs, window functions) to identify fraud patterns and user behavior.
- Built logistic regression and XGBoost classifiers in R, achieving 90%+ AUC and 0.78 balanced accuracy; engineered features and tuned thresholds to improve fraud recall.
- Visualized spending trends, category risks, and feature importance; combined SQL querying with R (ggplot2, caret) for interpretable model insights.

Credit Card Recommender System
- Engineered a KNN-based credit card recommendation system trained on a custom-built dataset scraped from 100+ cards across credit tiers.
- Automated feature extraction for cashback rate, APR range, annual fee, and bonus offers; cleaned and labeled data into a machine learning-ready format.
-Enabled users to input financial preferences and receive personalized card suggestions, simulating a fintech recommendation engine.

JOB EXPERIENCE

Front Desk Assistant, UCLA Housing
- Delivered high-quality customer service to students, staff, and visitors.
- Maintained and updated housing records and visitor logs, providing data-driven insights for administrative decision-making.